21561

Final Technical Report: Motion Analysis and Object Recognition For Autonomous Navigation

Contract: DACA76-89-C-0017

September, 1989 – March, 1991

United States Army Engineer Topographic Laboratories

Principal Investigators:

Edward M. Riseman and

Allen R. Hanson

Computer and Information Science

University of Massachusetts at Amherst

Amherst, MA 01003

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION *UNCLASSIFIED*	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT *Approved for public release; distribution is unlimited.*
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION *University of Massachusetts* / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION *U.S. Army Engineer Topographic Laboratories*
6c. ADDRESS (City, State, and ZIP Code) *Computer Science Department Amherst, Massachusetts 01003*	7b. ADDRESS (City, State, and ZIP Code) *Fort Belvoir, Virginia 22060-5546*
8a. NAME OF FUNDING/SPONSORING ORGANIZATION *Defense Advanced Research Projects Agency* / 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER *DACA76-89-C-0017*
8c. ADDRESS (City, State, and ZIP Code) *3701 N. Fairfax Drive Arlington, Virginia 22203-1714*	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. *0602301E*; PROJECT NO. *Program Code 0E20*; TASK NO.; WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Motion Analysis and Object Recognition for Autonomous Navigation -- Final Report

12. PERSONAL AUTHOR(S)
Edward M. Riseman and Allen R. Hanson

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM *9/89* TO *3/91*	*1992 April*	*40*

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
computer vision, autonomous vehicle navigation, landmark-based navigation, homing, planning, low level vision, structure from motion, pose recovery, stereo and motion,

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The research in computer vision described in this final report is directed towards the achievement of autonomous vehicle navigation using passive visual sensing. Navigation in both modeled and unknown environments has been investigated; in addition, research has been carried out in related areas of static image interpretation.

For a modeled environment, we have implemented a navigation system incorporating reactive planning, and based on the identification of known landmarks in the 3D scene. Also, robust algorithms have been demonstrated for the recovery of pose--the position and orientation of the camera--from model-matching between the image and known environment.

For an unknown environment, a navigation system has been demonstrated in which image-based homing is used to move between neighboring target locations. We have also developed and experimentally tested techniques for acquiring a detailed environmental model automatically, starting from an unknown or partially known environment. For a completely unknown environment, multiframe structure from motion algorithms have been developed which

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION *UNCLASSIFIED*
22a. NAME OF RESPONSIBLE INDIVIDUAL *Connie Gray*	22b. TELEPHONE (Include Area Code) *(703) 355-2733* / 22c. OFFICE SYMBOL *CEETL-RI*

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE

18. *model acquisition, kalman filtering, model-based matching, correspondence problem, robust statistics, obstacle detection, invariant image features, orientation statistics, shape from shading, smoothing, learning, perceptual organization, figure-ground discrimination, line grouping, camera calibration, high level vision, parallel architectures*

19. *use image sequences for the reconstruction of the camera motion and environmental structure. In a partially modeled environment, the combination of pose recovery with triangulation over image sequences yields a robust, accurate algorithm for incremental acquisition of a 3D scene model. Lastly, a new framework for obstacle detection from motion has been developed and demonstrated experimentally.*

In the area of static image interpretation and object recognition, research has been done on perceptual organization, invariant features, 3D reconstruction, and the automatic learning of strategies for object recognition. We have developed a new approach to distinguishing figure from ground, a prerequisite for obstacle detection, based on perceptual grouping techniques. Work has continued on the perceptual organization of image curves. The usefulness for object recognition of image features (approximately) invariant under motions of the camera was demonstrated, and the important theorem proven that in the general case no exactly invariant features exist. A new approach to the shape-from-shading problem was developed, leading to a fast robust algorithm for surface reconstruction from shading. Finally, general statistical methods for combining information on orientations were derived and shown to be useful for surface and line reconstruction.

Contents

1. Overview

Since the beginning of this contract, we have continued our work on algorithms for recovering motion and structure from image sequences, especially in their application to autonomous navigation. The robustness of motion algorithms developed at the University [ADI85] [SAW90b] [DUT90] has been evaluated experimentally in comparison with Horn's standard relative orientation algorithm [HOR90], using real images [SAW90c]. In addition, the robustness of stereo and motion algorithms has been analyzed theoretically [DUT91] [DUT90]. Work has also continued on camera calibration, which is essential for accurate motion and structure recovery.

New motion algorithms were developed, extending previous work through the use of simulated annealing [DUT90], through the combination of stereo with motion [BAL91], through integrating spatial and temporal constraints [SAW90a] [SAW90b], and through the use of a Kalman filter incorporating the effects of motion error [THO91] [OLI91b] [THO90]. In all cases, these algorithms were applied to real image sequences.

Robust algorithms for pose refinement, the determination of the camera position and orientation by model–matching in a known 3D environment, have been developed for use in our autonomous navigation project [KUM90b] [KUM90d]. The effects on pose refinement of uncertain knowledge of the camera parameters have been studied experimentally as well as theoretically [KUM90a] [KUM90c]. Also, techniques for learning a partially unmodelled environment using pose refinement have been experimentally tested [KUM90c].

Algorithms for establishing the correspondence between model and image, a prerequisite for pose refinment, have been developed by Beveridge [BEV91] [BEV90] [FEN90a]. Earlier work [BEV90] [FEN90a] solving the 2D-to-2D matching problem (where the estimated image projection of a landmark is matched directly to data, subject to rotation, translation, and scale) has been extended to the full 3D-to-2D matching problem [BEV91]. A comparative experimental study of these two approaches has been carried out [BEV91]. The results indicate that the 3D-to-2D matching system is more reliable, but slower, than the 2D-to-2D system, as expected; however, the 2D-to-2D system is rather robust in its own right in its ability to deal with perspective distortions.

The reactive planning software for controlling our mobile robot has been developed using a model of the second floor of our building [FEN91] [FEN90a] [FEN90b]. The locale modelling system was used to demonstrate the planning system. Software for the selection of environmental landmarks, to be used while the robot is executing an action, is part of the system. The steering system orients on detected environmental landmarks, and uses these to correct the robot's trajectory and verify its location during actions [FEN90b]. The system has been developed on the Sun 4 workstation, but the robot cannot perform action–level servoing in real time (for instance, moving the robot 20 feet takes about 3 minutes). With parallel architectures or other special purpose hardware, real–time navigation is achievable.

A new technique for robot navigation has been developed and implemented in our group, and is currently being explored by Pinette [HON91] [HON90] and Zhang [ZHA90]. This approach deals with the problem of automatic model acquisition of the three–dimensional environment. The robot navigates by using an image–based homing algorithm to move between neighboring target locations. A novel feature of this approach is an imaging system that acquires a compact, 360^0 representation of the environment.

Weiss [GRU91] [WEI90] [GRU90] has studied the problem of incrementally acquiring a surface representation for use in feedback control of a robot engaged in purposeful interaction with its environment. The incremental acquisition of a force–domain model for grasping is also described.

Sawhney has developed a potentially powerful framework for obstacle detection from motion [SAW90a]. In this approach, shallow structures in the image are segmented on the basis of consistent modelling of their image motion over time by affine transformations. Shallow structures are surfaces whose depth variation relative to their distance is small, and therefore can be represented by frontal plane surfaces parallel to the image plane. These ideas have been successfully tested using real images [SAW90a].

Work has continued also on static image interpretation and object recognition. Williams has used techniques of perceptual grouping to examine the difficult problem of distinguishing figure from ground [WIL90]. Solving this problem is a prerequisite for achieving obstacle avoidance. He has developed a system that is capable of distinguishing multiple planar figures in real images, despite the presence of multiple occlusions. Burns has presented a study of the variation in the appearance of 3D line features with respect to their 2D projection [BUR90]. 2D Features which display little variation are argued to be useful for object recognition. Also, he has proven that no view–invariant 2D feature exists in the general case of n points.

Draper has considered the problem of automatically learning object recognition strategies which are object-specific, from object descriptions and sets of interpreted training images [DRA90]. A separate recognition strategy is developed for every object in the domain. This work extends the knowledge-based approach by replacing the ad-hoc control heuristics of other systems with well-motivated control and classification decisions.

Collins has studied the problem of deriving 3D line and surface orientations from images by statistical methods. His work has been applied to determining line and plane orientations from stereo line pair correspondences [COL90b], and from a vanishing point analysis [COL90a].

A new theoretical understanding of shape from shading has been developed in [OLI91a] [OLI91c] [OLI91d] [OLI91e] [OLI90b] [OLI89]. These results imply that regularization is usually unnecessary for shape reconstruction, and will distort the recovered shape. Also, new constraints on the possible surface solutions corresponding to a shaded image were derived. These theoretical results have been incorporated in a new shape reconstruction algorithm that is simple, fast and robust—provably convergent (in many cases) to the correct surface.

A new local smoothing filter for curves and surfaces has been developed in [OLI90a], which combines the advantages of Fourier description and Gaussian smoothing. Unlike Gaussian filters, it does not exhibit the well–known problem of curve shrinkage.

Dolan has continued to work on the perceptual organization of image curves. In particular, he has been working on problems of redundant description, collateral grouping, and texture discrimination. He is designing a parallel implementation to run on the Connection Machine, and eventually on the Image Understanding Architecture [WEE89a]. He has also studied the problem of finding minimal length tree networks on the unit sphere [DOL91] [DOL89].

Snyder [SNY90a] [SNY90b] [SNY89] has achieved a general classification of the possible smoothness constraints for use in optical flow calculations, or surface interpolation, using the theory of group representations.

More work was done on the motion data set taken with the ALV at Martin Marietta [DUT89]. Three of the sequences were converted so that complete 3D information was available for the ground truth and the vehicle motion parameters.

Software is being developed for use on the Image Understanding Architecture (IUA)

[WEE89], developed by the University of Massachusetts in collaboration with Hughes. Dutta has implemented correlation matching on an instruction-level simulation of the IUA, which has been tested on a 64×64 image. It was found that for an image displacement of at most 5 pixels along any of the axes, the program ran in about 26 msec. The Nagin–Kohler segmentation algorithm [BEV89] and Boldt's line detection algorithm [BOL89], both developed in our group, are also in the process of being implemented.

A compiler of the language Apply has been implemented on the IUA simulator by Scudder. More work has been done on the Darpa Image Understanding Benchmark [WEE89b]. The low level portion was made more modular, and bugs were removed. Scudder is working on a database for the IUA which will be a successor to the sequential ISR (Intermediate Symbolic Representation) [BRO89]. Finally, a revised version of the ISR system for intermediate level vision has been implemented [DRA90b].

Much of the work described in this report has also been described in the annual report; an overview of this research appears in [RIS90].

2. Autonomous Robot Navigation

2.1 Landmark–based Navigation

Work on our autonomous navigation project utilized a model of the second floor of our building. The locale system for modelling the environment [FEN91] [FEN90a] [FEN90b] was used to demonstrate the planning system. In the locale system, the environment is represented as a graph which captures the key topological, geometric, and physical properties of the spatial entities making up the robot's world. This network describes space in terms of a hierarchical collection of locales: spatial entities representing objects, buildings, parking lots, free space, etc. A locale is a parcel of space which has semantic significance for the navigation problem.

The reactive planning software for controlling our mobile robot avoids the generation of detailed plans. Plans are "sketched" and modified in response to what is perceived as a result of each action. To begin with, each task given to the robot is decomposed depth first into a tree of less abstract subgoals. A sequence of subgoals from this tree is called a plan sketch because it is only partially developed, and generally will change as execution proceeds. The plan–and–monitor executive [FEN88] [FEN89] refines the plan–sketch hierarchy to fit the actual results of each action, in a procedure called *plan–level perceptual servoing.* Action begins as soon as the first subgoal in the plan sketch is a primitive (that is, capable of being directly executed by the vehicle).

A perceptual servoing cycle begins by analyzing what is known about the environment and what should be perceived from the current location of the agent. 3D entities, called landmarks, are selected from the environmental model on the basis of how distinctive they are, and what kind of information they offer the servoing procedure. Once these landmarks are selected their appearance is projected onto the image plane and matched to data in the image. These matches, along with the knowledge of the 3D locations of the landmarks, are used to compute the appropriate corrections to the plan sketch hierarchy.

Landmarks are selected on the basis that they will be easy to find using correlation. They are chosen to be distinctive surface patches defined by vertices separating regions of differing reflectance. For the current locale, all vertices from the model which are expected to be visible are collected. From these, the vertices associated with a reflectivity discontinuity above some threshold are returned as the selected landmarks.

Experiments demonstrated the use of correlation in matching a modelled landmark with its image for plan–level perceptual servoing. After matching, a version of Kumar's pose refinement algorithm [KUM90d] is used to calculate the robot's position. The experiments showed an accuracy of 0.15 feet over a depth range of 6 to 40 feet in our hallway sequence.

Perceptual servoing has been used at the action level [FEN91] [FEN90b] to orient the robot to detected environmental landmarks, which are also used to correct the robot's trajectory during the execution of a Move command. The system has been demonstrated successfully in our hallway environment; in these experiments, the robot moved in a straight line for distances of up to 40 feet with a tolerance of 0.3 inches [FEN91] [FEN90b]. On a Sun 4 workstation, the robot cannot perform action–level servoing in real time (for instance, moving the robot 20 feet takes about 3 minutes), but plans in the future include the IUA for real–time navigation.

2.2 Navigation via Homing

A new technique for robot navigation has been developed and implemented in our group, and is currently being explored by Pinette [HON91] [HON90] and Zhongfei Zhang [ZHA90]. This approach deals with the problem of automatic model acquisition of the three–dimensional environment. The problem is solved by representing the environment as a set of snapshots of the world taken at target locations. The robot navigates by using an image–based homing algorithm to move between neighboring target locations; explicit inference of three-dimensional structure from image data, though possible, is *not* required. A novel feature of this approach is an imaging system that acquires a compact, 360^0 representation of the environment.

This approach avoids the necessity for constructing a detailed 3D model of the robot's environment—a task that is often difficult and time–consuming in its own right. However, navigation is limited to a fixed set of target locations known to the robot. These locations are learned automatically by running the robot along a desired route, and having the system extract location signatures for a sequence of target locations on the route. After acquiring this model, the robot can navigate the route by successively homing to each of its target points.

Location signatures are obtained by using a novel and powerful imaging system to project a full 360° view of the world into a single image, which is then condensed into a compact, one-dimensional waveform representation, where significant changes in intensity provide landmark features. For homing to a nearby target location, the differences between the signatures of the robot's current and desired locations are used to compute incremental movements that take the robot closer to the target location. Homing can only be done locally; if the robot's current location is too far from the target location, the homing algorithm will fail because there will be too much distortion in images of the prominent landmarks common to both location signatures. Thus, large-scale navigation tasks are achieved by dividing them into a sequence of small-scale tasks that are solved by local homing. This complete system has been demonstrated on a mobile robot for a typical short-range navigation task.

In [ZHA90], a new method of matching landmarks between location signatures is employed. A symbolic representaion of the location signature is used first for qualitative matching; hypothesized matches are then verified quantitatively. In addition, a geometric model of the 3D environment is constructed. The rotation and translation between the current and target locations can be computed, as well as the distance to visible 3D landmarks. In an experiment carried out in our indoor hallway environment, the robot was able to home to the target location to within three inches and three degrees in orientation. This compares with an initital displacement of about two feet, and a turn of forty degrees in heading orientation.

2.3 Landmark–based Recovery of Structure and Motion

Over the last several years, we have developed algorithms for the estimation of camera location and orientation from a set of recognized landmarks appearing in the image. These algorithms for pose refinement have been applied as part of our system for autonomous robot navigation in a modelled environment [FEN91] [FEN90a]. In recent work, Kumar and Hanson have examined several of the factors affecting the robustness of pose determination for real image sequences. Also, they have applied their techniques to the problem of model extension.

An algorithm for recovering pose which is robust with respect to outliers was developed in [KUM90b] and [KUM90d]. The landmarks recognized and tracked in the image sequence are 3D lines or 3D points. Tracking is done using the line tracking algorithm of Williams [WIL88]. The algorithm can handle up to but less than 50% outliers, such as incorrect correspondences of the tracked landmarks. Several algorithms were tested on both indoor and outdoor scenes. It was found that determining the orientation and position of the camera simultaneously gave greater noise immunity than a sequential technique in which the orientation is computed first, followed by translation. An algorithm based on robust statistics which minimizes the *median* of the squared error was shown to be robust compared to the more usual approach in which the mean of the squared error is minimized. Several instances were reported in which the median–based approach was able to correctly determine the pose, despite wrong correspondences, while the standard technique failed. In [KUM90b], different algorithms based on robust statistics were compared.

In [KUM90a] [KUM90c], Kumar and Hanson studied the effect on pose refinement (and other related problems of 3D inference from 2D images) of errors in the image center and focal length. The analysis and conclusions are algorithm–independent. It has been determined that the image center can be in error by as much as 30 pixels for some standard imaging systems. Nevertheless, they demonstrated both theoretically and experimentally that for standard imaging systems, with a field of view significantly less than 90°, the recovered 3D location of the camera is insensitive to the position of the image center. However, the recovered camera orientation does depend linearly on the center offset. For instance, with a 24° field of view, and a 512×512 image, a 30 pixel center offset can cause a rotation error of about 1.5°. Experiments giving results in good agreement with the theoretical analysis were performed on two real image sequences [SAW90c] (these are described in detail in section 3).

It was demonstrated also [KUM90a] [KUM90c] that an incorrect estimate of the camera focal length only affects significantly the component of translation parallel to the optical axis of the camera. Experiments on the two real image sequences above confirmed this theoretical prediction.

Kumar has also been performing experiments on model extension [KUM90c]. Given a partial model of a scene and the results from the pose refinement algorithm for a sequence of images, the relative orientation between various image pairs is first computed. Then the depths of unmodelled points tracked over the sequence is computed by 'induced stereo'—by triangulation between image pairs using their previously determined relative orientation. The sensitivity of this depth–from–induced stereo algorithm to errors in the image center was also investigated.

In experiments using the first of the two real image sequences mentioned above (the box sequence), the 3D positions of the unmodelled points were recovered with an average error in depth of .25%. For the second sequence, the average error in depth was 1.3%. The error for the second case is larger than for the first in part due to the larger field of view (40° compared to 22°) which increases the sensitivity to image center offset. Also, for the larger field of view, radial lens distortion is expected to play a more important role. Given that there must be some error in the original 3D positions of landmarks, recovery of new 3D points to this accuracy is a surprising and dramatic result.

2.4 Model–Based Matching to Establish Correspondence for Pose Recovery

Finding the correct correspondence between the model (as projected in the image plane) and the image data is a prerequisite for pose recovery, and therefore crucial to our program of navigation using known landmarks. Beveridge [BEV91] [BEV90] [FEN90a] has developed algorithms based on local search which can determine this correspondence with some robustness. The features matched are straight lines.

Establishing correspondence is a combinatorial optimization problem. In the case of line matching, a match for each model line is sought from the set of data lines: often, more than one data line will be matched to a single model line (due to line fragmentation), and not all model/data lines will be matched. The matching problem is therefore difficult. Moreover, matching implicitly involves a second, subsidiary optimization problem: to measure the goodness of a hypothesized correspondence, the model must be rotated and translated to be in optimal spatial coincidence with the matching data. After this rotation and translation, the departure from perfect coincidence measures the goodness of the match.

In earlier work on landmark-based navigation [FEN90a] [BEV90], matching was done separately from and prior to 3D pose recovery. The 3D landmarks were projected into the 2D image plane, using an estimate of the robot's current pose; this 2D projection was then matched directly to the image data. To compensate for error in the pose estimate, it was assumed that a rotation, translation and scaling in the image plane were sufficient to bring the projected landmark into correct positioning. This assumption will be valid if the model is restricted in depth, and the estimate of the robot's position is reasonably good. It is invalid when an incorrect pose estimate reults in severe angular distortion of the projected model. So long as the assumption is valid matching reduces to a simple 2D problem. Moreover, Beveridge has shown that for this case the optimal transformations bringing a model into coincidence with data can be computed easily in closed–form [BEV90] [FEN90a], making possible a quick evaluation of a hypothesized corespondence.

To solve the combinatorial matching problem, Beveridge has introduced a novel variation of the local search approach: ***subset–convergent local search.*** In local search, locally optimal solutions are sought by incrementally modifying a hypothesized set of correspondences, beginning from random starting points. If this process is repeated many times, beginning at different starting points, the probability of determining the correct globally optimal correspondence approaches certainty. In subset–convergent local search, local search is done separately for different subsets of the model lines. The idea is that for a truly good match, a neighborhood search initiated for a subset should converge back to the good match. On the other hand, if the match is bad, then subsets of the match are probably incompatible, and searching with subsets often leads to a better match. Experiments show that this strategy works well [BEV90].

Beveridge has addressed the full 3D-to-2D matching problem in [BEV91]. In this case, the 3D model is matched to the image data; aligning the model to matching data can be done using the pose recovery algorithms of Kumar [KUM90b] [KUM90d]. This is more complicated than the 2D-to-2D approach previously considered, but it is robust against the perspective distortions produced by errors in the estimated 3D position of the sensor in world coordinates, which create difficulties for that approach. Local search is again used. In addition, Kumar's algorithm was modified using the Levenberg–Marquardt technique, for increased robustness. These two methods have been compared experimentally in [BEV91], using a hallway image sequence in which perspective effects on the projected model are appreciable, since the hallway model has large depth. It was found that the 2D-to-2D approach is more reliable than one might at first expect; in all cases, it improved the estimate of the robot's pose. The 3D-to-2D algorithm proved to be robust, always computing the essentially correct pose. The results also suggest that the additional cost of doing the full 3D-to-2D matching is not prohibitive. A hybrid algorithm, which blends 2D-to-2D and 3D-to-2D matching, may give a good compromise between speed and robustness, and is being investigated.

2.5 Obstacle Detection

Sawhney has developed a potentially powerful framework for obstacle detection [SAW90a]. In many man–made environments, obstacles in the path of a mobile robot can be characterized as *shallow*, i.e. they have relatively small extent in depth compared to the distance from the camera. In [Saw90a], it is demonstrated that shallow structures can be segmented using the property that their image motion is describable via an affine transformation. Structures emerge as shallow or non–shallow depending on whether they can be tracked consistently over time using the affine model of image motion. The tracking system, in turn, operates in a cycle of prediction (assuming affine motion), and generic model matching. This paper offers a new approach to tracking, rejecting heuristic assumptions on the motion or the similarity of tracked tokens, in favor of the consistency over time of a generic 3D model, namely, a shallow structure. Depths of the shallow structures can also be computed by this approach, without the intermediate step of explicit computation of the 3D motion parameters. Thus, the usually difficult problem of decomposing translation and rotation parameters in 3D motion can be avoided, while still recovering useful approximations to the depth of 3D surfaces.The system is useful for computing surface representations under both ego and independent–object motions.

An incremental algorithm is presented which works over a sequence of images captured by a camera undergoing smooth motion. Lines detected in the image are grouped into agregate structures which are hypothesized to form shallow structures. An initial location and motion are estimated for each hypothesized shallow structure, which are then updated in a prediction, matching and tracking cycle. Predictions are generated using the shallow structure model. Finally, true shallow structures are extracted from the set of hypothesized aggregate structures on the basis of consistent motion predictions and consistent depth of the structure. This gives a representation of the tracked shallow objects as frontal planar surfaces. The system was tested on an indoor image sequence with both ego–motion and independently moving objects, and worked well.

3. Motion and Structure Reconstruction

3.1 Robustness of General Motion Algorithms

In [SAW90c] two motion algorithms developed in our group [ADI85] [SAW90b] have been evaluated experimentally in comparison with Horn's standard relative orientation algorithm [HOR90], using image sequences obtained through a rotational motion of the camera in indoor scenes. Adiv's [ADI85] and Horn's algorithm are well–known general motion algorithms in which the scene structure is recovered from two image frames. The third algorithm, [SAW90b], is specialized to deal with rotational motion, and uses multiple frames. All the algorithms use point correspondences.

The first sequence consisted of a rectangular checkered box rotating around its body–axis. Ground truth was obtained by carefully measuring the coordinates of points on the box, and then using a pose estimation algorithm [KUM90d]. It was found that Adiv's algorithm [ADI85] compared well with Horn's [HOR90], as did the algorithm of Sawhney [SAW90b].

In the second sequence, the inside of a room was imaged. The camera was rigidly mounted on a robot arm, and the system rotated about an axis parallel to the optical axis. Thus, the camera motion consisted of rotation about an axis approximately perpendicular to the image plane, with simultaneous translation parallel to this plane. For this sequence, the general motion algorithms produced very unreliable and consistently biased depth estimates, as expected on theoretical grounds, while the third algorithm achieved good results. Due to the consistent bias, it is likely that even temporal integration will not improve the depths for the two–frame estimates.

In [DUT90], the robustness of two–frame motion algorithms is addressed theoretically. It is proven in an *algorithm–independent* way that small absolute errors in image displacements can cause significant errors in rotational motion parameters. Rotational errors of this magnitude can then produce large relative errors in the determination of environmental depth. Even if the motion parameters are known exactly, small errors in image displacements can still lead to large errors in depth for environmental points whose distance from the camera is greater than a few multiples of the total translation in depth of the camera.

Previously, the explanation for the lack of robustness in two–frame structure from motion algorithms had been sought in the context of specific algorithms, or specialized motion. The work described in [DUT90] is a first step towards a comprehensive, algorithm–independent study of the issues associated with the correspondence–based structure from motion problem. This analysis was extended to the case of binocular motion (the combined use of stereo and motion) in [DUT91], under the assumption of known motion.

3.2 New Motion Algorithms

Our development of new motion algorithms has extended previous work in several directions. Dutta [DUT90] has proposed and tested a new general motion, two–frame algorithm. Sawhney and Oliensis [SAW90b] have implemented an algorithm specialized for rotational motion, already mentioned above, which has better performance than two–frame algorithms in some situations. Stereo and motion have been integrated in an algorithm recovering motion and relative depth, developed by Balasubramanyam [BAL91]. Finally, Thomas and Oliensis [THO91] [OLI91b] [THO90] have developed a technique for recursively determining structure from multi–frame image sequences using a Kalman filter, which compensates for the motion error made by two–frame algorithms.

Dutta has developed a new algorithm for recovering depth from general motion, which he uses to refine the results obtained by other algorithms (such as Adiv's [ADI85]). A new objective function is proposed, which essentially tries to minimize the difference in the depths computed by the x and y–components of the image displacements. This function has the nice property that it gives an estimate of the average reliability of the depth measurements. A fast simulated annealing algorithm is used to minimize this objective function.

In [SAW90b] is presented a new technique for reconstructing the 3D structure and motion of a scene undergoing relative rotational motion with respect to the camera. Given image correspondences of point features tracked over many frames, a two–stage technique for reconstruction is presented. First, a grouping algorithm is developed which exploits spatio–temporal constraints of the common motion to achieve a *reliable* description of discrete point correspondences as curved trajectories in the image plane (general conics in the case of rotational motion). Image trajectories are grouped on the basis of proximity in the image (presumed to imply proximity in space), and goodness of combined fit for the given motion model. Secondly, the 3D motion and structure are solved for in closed form from the computed image trajectories. The closed form solution, valid for *perspective projection*, is a new result. The algorithm has been demonstrated on real image sequences with good results, as described above.

It is argued in this paper that both spatial and temporal context should be exploited for reliable structure recovery. Thus, a single 3D point trajectory often yielded very ambiguous reconstructions, in contrast with a combined fit to spatially–grouped trajectories. Also, for a motion sequence with motion parallel to the image plane, structure recovered using the standard, temporally–limited, two–frame motion algorithms was wrong and *biased*. This problem is not likely to be cured by combining many different two–frame estimates, for instance using a Kalman filter, because of the consistent bias. On the other hand, Sawhney's algorithm, which integrates information over time using many frames simultaneously, produced the correct result. Techniques for recovering structure from rotational motion may be useful for automatic model acquisition in industrial settings by inducing rotational motion of the sensor.

In [BAL91], Balasubramanyam addresses the problem of recovering structure and motion using a *binocular* camera system. Combining stereo with motion in this way gives increased robustness of depth recovery, since the depths are determined redundantly at every time step. It is demonstrated that a vector encoding 3D information—the *p–field*—can be derived directly from purely image measurable quantities, namely, the optic flow and stereo disparity. For each image point, the p–field is parallel to the real instantaneous 3D velocity vector for the corresponding 3D point, scaled by the depth of this point.

Balasubramanyam also examines the behavior of the p–field for specific motions (purely rotational, purely translational), as well as for general motion, and gives experimental results on real and synthetic binocular image sequences. Finally, possible applications are discussed. For instance, since the p–field is a scaled replica of the real 3D velocity vector, it seems more appropriate to impose smoothness on this vector, rather than on optical flow, in deriving smoothed motion fields. It may also provide a useful framework for occlusion detection.

In [THO91] [OLI91b] [THO90] a new algorithm for recursively determining structure from multi–frame image sequences using a Kalman filter was implemented. Input for the algorithm consists of point correspondences tracked over many image frames. This work incorporates the insight, cited above [DUT90], that small errors in recovering motion using a two–frame algorithm can cause large errors in structure determinination. Thus, in recursively combining structure estimates from multiple frame pairs of an image sequence, it is important to maintain a record of the effects due to motion error, and to compensate for these effects. This was not done in previous recursive, multi–frame algorithms. The new algorithm of [OLI91b] [THO91], which does compensate for the effects of motion error, has achieved good results for general motion on synthetic and real images sequences.

The algorithm is based on the observation that errors in the motion produce *cross–correlations* in the structure errors between different 3D points. Conversely, these correlations are the record of the motion error. Thus, to explicitly incorporate motion error in a recursive algorithm, a record of the correlations in the structure errors must be maintained and updated.

Horn's relative orientation algorithm [HOR90] is used to provide two–frame structure estimates. For this algorithm, a somewhat complex error analysis is used to estimate the expected structure errors, including the cross–correlations. The fusing of the new structure estimate with the old is done using a standard Kalman filter, but with the cross–correlations taken into account. Typically, a complete computation for fourteen points tracked over fifteen time steps required thirty minutes on a TI explorer. The results on synthetic images show that the structure estimates improve over time as expected. For one of the real image sequences, the improvement is dramatic after only four frames; this can be explained as due to the algorithm having corrrectly combined successive measurements to obtain an effectively wider baseline.

3.3 Incremental Visual Modeling for Feedback Control

In [GRU91] and [GRU90], the incremental construction of geometrical and force-domain models for closed-loop sensing and control from visual images is described. The force model consists of a two-dimensional surface in a six-dimensional wrench space. It is computationally desirable to have multiple resolution force domain models which use solutions at a coarse level to provide initial conditions for solutions at a higher resolution. Reasoning in the force domain is facilitated by constructing multiple resolution models. Spatial frequency encoded models based on Fourier coefficients are explored in these two papers.

Also, it is shown how incomplete surface data derived from a sequence of visual images can be used to incrementally construct a local geometric model, represented by planar patches. Results are presented for a sequence of images derived from a test object (soda can). It is then shown how this model can be mapped to a global wrench space model encoded by the Fourier coefficients.

The incremental acquisition of sparse geometric information (position, orientation, and curvature) was also studied in [WEI90]. Occluding contours, surface markings, and creases are tracked over multiple views. 3D polygonal curves and triangular surface patches are computed from known camera motion. The sectional curvature in the viewing direction can also be approximately computed. Results were presented for a rotating cylinder and Rubik's cube.

4. Object Recognition: Static Image Understanding

4.1 View Variation of Image Features

Burns has presented a study of the variation in the appearance of line and 2D features with respect to the view [BUR90]. He argues that if an image feature is to be useful in object recognition, the feature should vary by a small amount over useful views. His paper examines feature variation under a restricted but useful class of camera views—those for which the camera is sufficiently far from the viewed object that *weak perspective* is an appropriate imaging model. For such views, the depth variation on the object should be less than about one–tenth of the distance to the camera. It is shown that some simple features vary little over most of the view sphere and are thus potentially useful for object recognition—for example, the angle between two 3D line segments when this angle is small.

Also, he considers features that are strictly *invariant* under variations of the view. View–invariant features are obviously ideal for object recognition. For general point sets, however, Burns has proven that *no invariant feature exists under perspective projection* [BUR90]. This is a new result. It holds also for orthographic and weak perspective projection. There do exist special–case view invariants of practical importance—image features that are only view–invariant for special configurations of 3D points. Burns provides a classification of such features under weak perspective, and derives some new invariants that have not appeared previously in the recognition literature.

4.2 Determining Line and Plane Orientations by Statististical Methods

Collins has studied the problem of deriving 3D line and surface orientations from images by statistical methods [COL90a] [COL90b]. Since the orientation of a 3D line or planar surface can be specified by a unit vector, he examined the properties of probability distributions of such vectors on the unit sphere, and statistical inferencing techniques over such distributions.

His work can be applied to determining line and plane orientations from stereo line pair correspondences [COL90b], or from a vanishing point analysis [COL90a]. In both cases, the problem can be posed as follows: given a set of orientation vectors approximately determined from the image, determine the orientation vector that is most nearly perpendicular to all of these. For instance, for a stereo line pair correspondence, the *projection plane* is defined for each image as the plane spanning the camera focal point and the 2D image line. Since the 3D line must lie in the projection planes for both images, its orientation vector is determined as being perpendicular to the normal vectors for these planes.

The statistical problem is to determine accurately the perpendicular unit vector given a set of approximately coplanar vectors obtained with some uncertainty, and to estimate confidence regions for this unit vector. In Collins' approach, the measured vectors are assumed to be distributed in accordance with a Bingham's distribution, which is essentially a Gaussian distribution restricted to the sphere. To determine a best estimate of the perpendicular unit vector, and confidence regions for this estimate, the parameters of the Bingham distribution must be estimated from the sample of measured vectors. Experimental results on real images for the problem of determining planar orientations from a vanishing point analysis are presented using this technique, and also a non–parametric estimation procedure. The two techniques gives results in good agreement with each other and with ground truth. However, this approach is complex and computationally expensive.

Collins also shows how to derive an easily computable approximation for the Bingham confidence regions, by computing the least–square best–fit plane for the sample measurements on the sphere. Finally, he shows how to solve the stereo problem using a somewhat different approximating technique. Experimental results on real images achieved accurate recovery of ground truth for this problem domain also.

4.3 High Level Vision: Learning Object Recognition Strategies

Draper has considered the problem of automatically learning object recognition strategies that are object-specific, from object descriptions and sets of interpreted training images [DRA90]. A separate recognition strategy is developed for every object in the domain. The goal of each recognition strategy is to identify any and all instances of the object in an image, and give the 3D position (relative to the camera) of each instance. The goal of the learning process is to build a strategy that minimizes the expected cost (in our case computational cost) of recognition subject to accuracy constraints imposed by the user.

In this work, object recognition is modeled as a process of applying visual knowledge sources to hypotheses, where the knowledge sources are standard image understanding strategies, such as 2D $\rightarrow$ 3D point matching, vanishing point analysis, and straight line extraction. Hypotheses are intermediate–level statements about the image and/or 3D world, and can occur at many levels of abstraction. Examples include straight line segments, 3D orientation vectors, and volumes. At each step in the recognition process, a knowledge source is applied to one or more hypotheses. The result is either a new hypothesis or a discrete evidence value reflecting the quality of the original hypotheses.

Recognition strategies are represented by recognition graphs, which are similar in many ways to decision trees. Unlike decision trees, however, recognition graphs direct hypothesis creation as well as hypothesis classification or verification. Object–specific strategies are learned in a two step process. The first step involves learning which hypotheses should be generated. The second learns how to verify them efficiently.

This work extends the knowledge-based approach to image understanding by replacing the ad-hoc control heuristics of other systems with well-motivated control and classification decisions. The user, instead of supplying heuristics in the form of if–then rules or confidence functions, specifies accuracy requirements. The system selects knowledge sources that minimize the expected cost of recognition while achieving the specified accuracy.

4.4 Figural Completion and Figure–Ground Separation

Williams has used techniques of perceptual grouping to examine the difficult problem of distinguishing figure from ground [WIL90]. Solving this problem is a prerequisite for achieving obstacle avoidance. He has developed a system that is capable of distinguishing multiple planar figures in real images, despite the presence of multiple occlusions. The system selects the optimal consistent interpretation of the different figures, hypothesizing possible figural completions (i.e., hidden lines) when there is occlusion, and grouping figure boundaries into complete surface interpretations. The system is able to solve relatively complex perceptual problems such as the construction of an illusory triangle for the Kanizsa Triangle.

The system operates in two stages. In the *problem posing* stage, image evidence is collected and incorporated in a graph, called the contour graph. This graph includes *image lines* recovered using a straight line detection algorithm, *virtual lines* which are hypothetical joins between collinear image lines (these may have been missed by the line detector due to occlusion or low image contrast), *endpoints* of the image lines, *corners* between image lines, and finally *crossings* where image or virtual lines intersect.

In the second *problem solving* stage, the features created in the first stage are organized into the optimal figural completions, using integer linear programming. The lines are grouped into surface boundaries. Virtual lines are identified as occluded contours, or as visible lines below the contrast threshold of the line extraction algorithm, or rejected as not corresponding to real scene boundaries. The signs of occlusion at crossings are determined, as well as the occlusion–ordering of the reconstructed surfaces.

The linear program incorporates physical constraints specifying the mechanics of occlusion and figural completion. An example of such a constraint is the requirement that every occluding contour must have one of two signs of occlusion. Also, since the projections of complete surface boundaries are closed contours, all occluding contours must be contained in cycles in the surface boundary graph. Imposing such constraints leads to many *feasible solutions*, each a physically possible interpretation of the image. Additional *preference* criteria, such as a preference for figures which are convex at corners, are added by means of a linear objective function, which is to be minimized subject to the physical constraints. In experiments carried out for a Colorforms image domain (a 2D planar world), the minima of the objective function computed by the integer linear program corresponded to the human–preferred interpretations.

4.5 Shape from Shading

In several recent papers, Oliensis [OLI91a] [OLI91c] [OLI91d] [OLI91e] [OLI90b] [OLI89] has presented a new theoretical understanding of shape from shading. Although this problem has traditionally been considered to be ill–posed, Oliensis has shown that the surface solution corresponding to a shaded image is often strongly constrained. Moreover, for images of general smooth objects wholly contained in the field of view, and for illumination from (or symmetric around) the camera direction, he has proven that shape is ***uniquely*** determined by shading [OLI91c] [OLI89]. A fortiori, it is well–posed under these conditions. This is the first uniqueness result valid for images of general surfaces.

The case of general illumination direction is analyzed in [OLI91d] [OLI91e]. Several constraints on the potential surface solutions are derived, and it is argued that for a typical image, shading determines shape essentially up to a finite ambiguity. More conjectural arguments suggest that in fact shape is often determined with little ambiguity. However, although shape from shading is typically well–constrained, this is not always true: the strength of the constraint on the surface solution depends on the image. For some images, the reconstruction is uniquely determined. On the other hand, an explicit example is discussed where the surface reconstruction is uniquely determined over most of the image, but infinitely ambiguous within a small image region. For this image, shape from shading is a ***partially*** well–posed problem. It is argued that such ill–posed regions can occur frequently near the image boundary, but typically are small fractions of the image. A practical consequence is that the image data near the boundary should be given less weight in a shape reconstruction algorithm. Ideally, the surface should be reconstructed from the interior of the image outwards—otherwise, the potential instability of the boundary surface solution may propagate errors inwards, even if the surface in this region is actually well–determined.

[OLI90d] [OLI90e] also contain an analysis of the constraint on shape imposed by the image of the occluding boundary. This boundary has traditionally been thought to give a strong constraint, since the surface orientation is determined along it. However, it is proven in these papers that this is false, and that the surface reconstruction near the occluding boundary is actually more ambiguous than it is in the neighborhood of an interior image line.

Oliensis has also studied 'impossible images'—those with no corresponding smooth surface solution. In [OLI89] [OLI91c], again for illumination from the camera direction and an imaged object within the field of view, it is shown that a general image can be converted into an effectively impossible one by a small perturbation of its intensities. Thus for almost all such images, (i.e., intensity functions $I(x, y)$), effectively no smooth solution to shape from shading exists. However, non–smooth solutions will always exist [OLI91a].

The new constraints on the surface solutions for shape from shading derived in the work above have been incorporated in a shape reconstruction algorithm [OLI91a] that is simple, fast, and robust—provably convergent (in many cases) to the correct surface. This algorithm does not use regularization unlike previous ones. This paper also contains a simple uniqueness proof for shape from shading, and gives an explicit representation for the surface corresponding to an image. Also, a new local algorithm for shape from shading was proposed in [OLI90b].

4.6 Curve Smoothing without Shrinkage

Oliensis has derived a simple local smoothing filter for image or space curves, which combines the advantages of Gaussian smoothing and Fourier curve description [OLI90a]. Unlike Gaussian filters, it has no shrinkage problem. Repeated application of the filter does not yield a curve smaller than the original, but simply reproduces the result that would have been obtained by a single application at the largest scale. Unlike Fourier description, the filter is local in space, i.e., it has a limited spatial width. Thus, it can conveniently be applied by convolving in space. Also, the result of smoothing a curve segment does not depend on whether or not it is embedded in a longer curve. The method yields as a byproduct a compact description of the smoothed curve. Experimental results are presented for open as well as closed 2D contours.

Oliensis traces curve shrinkage under Gaussian smoothing to the fact that a Gaussian filter reduces the amplitudes of *all* spatial frequency components in the signal to which it is applied. He demonstrates that smoothing without shrinkage can be obtained using a hard cutoff low pass filter, which is equivalent to the use of Fourier descriptors. However, such a filter essentially has infinite spatial width. Therefore, a slight modification of this filter is proposed, which is shown to have limited spatial width. Rather than a hard high frequency cutoff, the modified filter has a slightly smoothed rolloff in its response to high frequencies.

4.7 A Complete Classification of Smoothness Constraints

Gradient–based approaches to the computation of optical flow often use a minimization technique incorporating a smoothness constraint on the optical flow field. Smoothness constraints are also of interest in surface interpolation, where they are known as "performance functions." All known smoothness constraints used to compute optical flow have a subtle property, namely that they do not mix derivatives of different components of the optical flow field. Snyder [SNY90a] [SYN90b] presents an analysis of smoothness constraints which do not satisfy this "decoupled" property, but rather in which derivatives of different components of the flow can interact. By using the single, natural assumption that a smoothness constraint should not change form under transformations to different Cartesian coordinate systems, Snyder determines a *complete* list of all possible invariant smoothness constraints of type (p, q), by which it is meant that they are quadratic in p^{th} derivatives of the optical flow field, and in q^{th} derivatives of the grey level image intensity function. This is done explicitly for the values $0 \leq p, q \leq 2$. All of these smoothness constraints, excepting those linear combinations which are decoupled, are new. In addition, Snyder finds all invariant "performance measures," used in surface interpolation, when the performance measure is quadratic in no higher than fourth derivatives of the objective function. These results are based on the representation theory of the group of Euclidean motions in the image plane. They extend earlier work on uncoupled constraints which will appear in [SNY89].

5. References

[ADI85] Adiv, G., "Determining Three-Dimensional Motion and Structure From Optical Flow Generated By Several Moving Objects", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-7, No. 4, July 1985, pp. 384-401.

[BAL91] Balasubramanyam, P., and Snyder, M., "The P–Field: A Computational Model for Binocular Motion Processing", *Proc. IEEE Computer Vision and Pattern Recognition*, Lahaina, Maui, Hawaii, June 1991, pp. 115-120.

[BEV91] Beveridge, J. Ross, and Riseman, E.M., "Hallway Navigation in Perspective", *AAAI Fall Symposium, Sensory Aspects of Robotic Intelligence*, Asilomar, CA, November 14-17, 1991, to appear.

[BEV90] Beveridge, J. Ross, Weiss, R., and Riseman, E.M., "Combinatorial Optimization Applied to Variable Scale 2D Model Matching," in *Proc. of the IEEE International Conference on Pattern Recognition*, Atlantic City, New Jersey, June 1990, pp. 18-23.

[BEV89] Beveridge, J. R., Griffith, J., Kohler, R. R., Hanson, A. R., and Riseman, E. M., "Segmenting Images Using Localized Histograms and Region Merging", *International Journal of Computer Vision*, Vol. 2, No. 3, pp. 311–347, 1989.

[BOL89] Boldt, M., Weiss, R., and Riseman, E. M., 'Token Based Extraction of Straight Lines", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 19, No. 6, pp. 1581–1594, 1989.

[BRO89] Brolio, J., Draper, B., Beveridge, J. R., and Hanson, A., "ISR: A Database for Symbolic Processing in Computer Vision", *IEEE Computer*, Vol. 22, No. 12, pp. 22–30, 1989.

[BUR90] Burns, J.B., Weiss, R., and Riseman, E., "View Variation of Point Set and Line Segment Features", *Proc. of the DARPA Image Understanding Workshop*, Pittsburgh, PA, September 1990, pp. 650-659. Also COINS Technical Report 90-84, September 1990.

[COL90a] Collins, R., and Weiss, R., "Vanishing Point Calculation as a Statistical Inference on the Unit Sphere", *Proc. IEEE International Conference on Computer Vision*, Osaka, Japan, December 1990, pp. 400-405.

[COL90b] Collins, R., and Weiss, R., "Deriving Line and Surface Orientation by Statistical Methods", *Proc. of the DARPA Image Understanding Workshop*, Pittsburgh, PA, September 1990, pp. 433-438. Also COINS Technical Report 90-102.

[DOL91] Dolan, J., Weiss, R., Smith, J. M., "Minimal Length Tree Networks on the Unit Sphere", to appear in *Journal Of Operations Research*, special issue on topological network design, 1991.

[DOL89] Dolan, J., Weiss, R., Smith, J. M., "Minimal Length Tree Networks on the Unit Sphere", Computer & Information Science Technical Report 89-105, University of Massachusetts at Amherst, October 1989.

[DRA91]Draper, B., and Hanson, A. "An Example of Learning in Knowledge-Directed Vision", *The 7th Scandinavian Conference on Image Analysis*, Aalborg University, Denmark, August 13-16, 1991.

[DRA90a] Draper, B., and Riseman, E., "Learning 3D Object Recognition Strategies", *Proc. IEEE International Conference on Computer Vision*, Osaka, Japan, December 1990, pp. 320-324.

[DRA90b] Draper, B., Beveridge, J.R., Brolio, J. Hanson, A., Heller, R., and Williams, L., "ISR2 User's Guide", University of Massachusetts TR 90-52.

[DUT91] Dutta, R., and Snyder, M.A, "Robustness of Structure From Binocular Known Motion", to appear in *Proc. IEEE Workshop on Visual Motion*, Princeton, NJ, October 6-9, 1991.

[DUT90] Dutta, R., and Snyder, M.A., "Robustness of Correspondence-Based Structure From Motion", *Proc. IEEE International Conference on Computer Vision*, Osaka, Japan, December 1990, pp. 106-111. Also *Proc. of the DARPA Image Understanding Workshop*, Pittsburgh, PA, September 1990, pp. 433-438.

[DUT89] Dutta, R., Manmatha, R., Williams, L. R., and Riseman, E. M., "A Data Set for Quantitative Motion Analysis", *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA, June 1989.

[FEN91]Fennema, C., "Interweaving Reason, Action and Perception," University of Massachusetts Technical Report 91-56, 1991.

[FEN90a] Fennema, C., Hanson, A., Riseman, E., Beveridge, J.R., and Kumar, R., "Model-Directed Mobile Robot Navigation", *IEEE Transactions on Sytems, Man and Cybernetics*, special issue on unmanned systems and vehicles, Vol. 20, No. 6, December 1990.

[FEN90b] Fennema, C., and Hanson, A., "Experiments in Autonomous Navigation", *IEEE International Conference on Pattern Recognition*, Atlantic City, NJ, June 1990, Vol. 1, pp. 24-31. Similar versions of this paper also appeared in *IEEE International Workshop on Intelligent Motion Control*, Istanbul, Turkey, August 1990, Vol. 1, IP-29-37, and *Proc. of the DARPA Image Understanding Workshop*, Pittsburgh, PA, September 1990, pp. 772-781.

[FEN89] Fennema, C., Hanson, A., and Riseman, E., "Towards Autonomous Mobile Robot Navigation", *Proc. of the DARPA Image Understanding Workshop*, Palo Alto, CA, May 1989, pp. 219-231.

[FEN88] Fennema, C., Riseman, E., and Hanson, A., "Planning With Perceptual Milestones To Control Uncertainty in Robot Navigation", *Proc. of SPIE International Society for Photographic and Industrial Engineering*, Cambridge, MA, November 1988.

[GRU91] Grupen, R. and Weiss, R., "Force Domain Models for Multifingered Grasp Control", submitted to *International Conference on Robotics and Automation*, 1991.

[GRU90] Grupen, R., Weiss, R., and Oskard, D., "Grasp Oriented Sensing and Control", *Proc. SPIE Applications in Optical Science and Engineering*, Hynes Convention Center, Boston, MA, November 1990.

[HON91] Hong, J., Tan, X., Pinette, B., Weiss, R., and Riseman, E., "Image-Based Homing", *IEEE Robotics and Automation Conference*, Sacremento, CA, April 1991, to appear. Also COINS TR 91-15.

[HON90] Hong, J., Tan, X., Pinette, B., Weiss, R., and Riseman, E., "Image–Based Navigation Using 360° Views", *Proc. of the DARPA Image Understanding Workshop*, Pittsburgh, PA, September 1990, pp. 782-791.

[HOR90] B. K. P. Horn, "Relative Orientation", *International Journal of Computer Vision*, Vol. 4, pp. 59–78, 1990.

[KUM90a] Kumar, R., and Hanson, A., "Sensitivity of the Pose Refinement Problem to Accurate Estimation of Camera Parameters", *Proc. of the IEEE International Conference on Computer*

Vision, Osaka, Japan, December 1990, pp. 365-369.

[KUM90b] Kumar, R., and Hanson, A., "Analysis of Different Robust Methods for Pose Refinement", *IEEE International Workshop on Robust Computer Vision*, Seattle WA, October 1990, pp. 167-182.

[KUM90c] Kumar, R., and Hanson, A., "Pose Refinement: Application to Model Extension and Sensitivity to Camera Parameters", *Proc. of the DARPA Image Understanding Workshop*, Pittsburgh, PA, September 1990, pp. 660-669.

[KUM90d] Kumar, R., and Hanson, A., "Robust Estimation of Camera Location and Orientation From Noisy Data With Outliers ", Computer & Information Science Technical Report 89-120, University of Massachusetts at Amherst, December 1989.

[OLI91a] Oliensis, J., "Direct Method for Reconstructing Shape from Shading," to appear in *SPIE Conference 1570 Geometric Methods in Computer Vision,* San Diego, California, July 1991.

[OLI91b] Oliensis, J., and Thomas, J.I., "Incorporating Motion Error in Multiframe Structure From Motion", to appear in *Proc. IEEE Workshop on Visual Motin*, Princeton, NJ, October 6-9, 1991.

[OLI91c] Oliensis, J., "Uniqueness in Shape From Shading", *International Journal of Computer Vision, 6:2*, pp. 75-104, 1991.

[OLI91d] Oliensis, J., "Shape From Shading as a Partially Well-Constrained Problem", to appear in *Computer Vision, Graphics, and Image Processing: Image Understanding*, 1991.

[OLI91e] Oliensis, J., "Shape From Shading as a Partially Well-Constrained Problem", *Proc.*

IEEE Computer Vision and Pattern Recognition, Lahaina, Maui, Hawaii, June 1991, pp. 559-564.

[OLI90a] Oliensis, J., "Local Reproducible Curve Smoothing Without Shrinkage", submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence*, December 1990.

[OLI90b] Oliensis, J., "New Results in Shape From Shading", *Proc. DARPA Image Understanding Workshop*, Pittsburgh, PA, September 1990, pp. 145-153.

[OLI89] Oliensis, J., "Uniqueness and Existence in Shape From Shading", *IEEE International Conference on Pattern Recognition*, Atlantic City, NJ, June 1990, Vol. 1, pp. 341-345.

[RIS90] Riseman, E., and Hanson, A., "Progress in Computer Vision at the University of Massachusetts", *Proc. of the DARPA Image Understanding Workshop*, Pittsburgh, PA, September 1990, pp. 86-96. Also COINS TR 90-100.

[SAW90a] Sawhney, H., and Hanson, A., "Identification and 3D Description of 'Shallow' Environmental Structure in a Sequence of Images", *Proc. IEEE Computer Vision and Pattern Recognition*, Lahaina, Maui, Hawaii, June 1991, pp. 179-185.

[SAW90b] Sawhney, H., Oliensis, J., and Hanson, A., "Image Description and 3D Interpretation From Image Trajectories", *IEEE International Conference on Computer Vision*, Osaka, Japan, December 1990, pp. 494-498.

[SAW90c] Sawhney, H., and Hanson, A., "Comparative Results of Some Motion Algorithms on Real Image Sequences", *Proc. of the DARPA Image Understanding Workshop*, Pittsburgh, PA, September 1990, pp. 307-313.

[SNY90a] Snyder, M. "The Mathematical Foundations of Smoothness Constraints: A New Class of Coupled Constraints," *Proc. of the DARPA Image Understanding Workshop*, Pittsburgh, PA, September 1990, pp. 154-161.

[SNY90b] Snyder, M. "The Mathematical Foundations of Smoothness Constraints: A Complete Enumeration of A New Class of Coupled Constraints", 1990.

[SNY89] Snyder, M. "On the Mathematical Foundations of Smoothness Constraints for the Determination of Optical Flow and for Surface Reconstruction", to appear in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991.

[THO91]Thomas, J.I., and Oliensis, J., "Incorporating Motion Error in Multiframe Structure From Motion", *The 7th Scandinavian Conference on Image Analysis*, Aalborg University, Denmark, August 13-16, 1991.

[THO90] Thomas, J.I., and Oliensis, J., "Fusing Structure by Kalman Filtering", Computer Science Technical Report 90-93, University of Massachusetts at Amherst, May 1990.

[WEE89a] Weems, C., Levitan, S., Hanson, A., Riseman, E., Shu, D., Nash, J. G., "The Image Understanding Architecture", *International Journal of Computer Vision*, Vol. 2, No. 3, pp. 251-282, 1989.

[WEE89b] Weems, C., Hanson, A., Riseman, E., "A Report on the Results of the DARPA Integrated Image Understanding Benchmark Exercise", *Proc. of the DARPA Image Understanding Workshop*, Palo Alto, CA, May 1989, pp. 165–192.

[WEI90] Weiss, R., Grupen, R., Oskard, D., "Visual Modeling for Feedback Control," 1990.

[WIL90] Williams, L., "Perceptual Organization of Occluding Contours", *IEEE International Conference on Computer Vision*, Osaka, Japan, December 1990, pp. 133-137. Also *Proc. of the*

DARPA Image Understanding Workshop, Pittsburgh, PA, September 1990, pp. 639-649.

[WIL88] Williams, L., and Hanson, A., "Translating Optical Flow Into Token Matches", *Proc. DARPA Image Understanding Workshop*, Cambridge, MA, April 1988.

[ZHA90] Zhang, Z., Weiss, R., and Riseman, E., "Segment-Based Matching For Visual Navigation", *Proc. IEEE Computer Vision and Pattern Recognition*, Lahaina, Maui, Hawaii, June 1991, pp. 742-743.

2/5/11